

SEQUENCE LISTING

<110> K .U. Leuven Research & Development

Debyser, Zeger

De Clercq, Erik

Cherepanov, Peter

Pluymers, Wim

<120> A synthetic gene for expression of a retroviral protein
with wild type activity in eukaryotic cells

<130> K1291-PCT

<140>

<141>

<150> EP99201306.0

<151> 1999-04-26

<150> EP00200171.7

<151> 2000-01-18

<160> 2

<170> PatentIn Ver. 2.1

<210> 1

<211> 930

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:synthetic gene
expressing HIV integrase

<220>

<221> CDS

<222> (27)..(899)

<220>

<221> misc_signal

<222> (24)..(30)

<223> Kozak sequence

<400> 1

atcactagca acctcaaaca gacacc atg gga ttc ctg gac ggc att gac aag 53
Met Gly Phe Leu Asp Gly Ile Asp Lys

1

5

gct cag gag gag cac gag aag tac cac tcg aat tgg cgg gcc atg gcc 101
 Ala Gln Glu Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala Met Ala
 10 15 20 25

tcc gac ttc aac ctg cca ccc gtc gtc gct aag gag atc gtt gct agc 149
 Ser Asp Phe Asn Leu Pro Pro Val Val Ala Lys Glu Ile Val Ala Ser
 30 35 40

tgc gac aag tgc cag ctg aaa ggc gag gct atg cac ggg cag gtt gat 197
 Cys Asp Lys Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln Val Asp
 45 50 55

tgc tct ccc ggc atc tgg cag ctc gac tgt act cac ctg gag ggc aag 245
 Cys Ser Pro Gly Ile Trp Gln Leu Asp Cys Thr His Leu Glu Gly Lys
 60 65 70

gtc atc ctg gtc gcc gtg cac gtg gcc tct ggt tac atc gag gct gag 293
 Val Ile Leu Val Ala Val His Val Ala Ser Gly Tyr Ile Glu Ala Glu
 75 80 85

gtc atc cct gca gag act ggc cag gag act gcc tat ttc ctg ctg aaa 341
 Val Ile Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Leu Leu Lys
 90 95 100 105

ctg gcc ggc cgg tgg cct gtg aag aca gtg cac aca gat aac ggc tcc 389
 Leu Ala Gly Arg Trp Pro Val Lys Thr Val His Thr Asp Asn Gly Ser
 110 115 120

aac ttc acc tcc acc act gtg aag gct gcc tgc tgg tgg gct ggg atc 437
 Asn Phe Thr Ser Thr Thr Val Lys Ala Ala Cys Trp Trp Ala Gly Ile
 125 130 135

aag cag gag ttc ggg atc ccc tat aac cca cag tct cag ggc gtg atc 485
 Lys Gln Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly Val Ile
 140 145 150

gaa tcc atg aac aag gag ctg aag aag atc atc ggc cag gtt cgg gac 533
 Glu Ser Met Asn Lys Glu Leu Lys Lys Ile Ile Gly Gln Val Arg Asp
 155 160 165

cag gca gag cac ctg aag act gca gtg cag atg gcc gtg ttc atc cac 581
 Gln Ala Glu His Leu Lys Thr Ala Val Gln Met Ala Val Phe Ile His
 170 175 180 185

aac ttc aag cga aag ggc ggc atc ggt ggc tac tca gcc ggc gag cgg 629
 Asn Phe Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly Glu Arg
 190 195 200

atc gtg gac atc atc gcc act gac atc cag acc aaa gag ctg cag aag 677
 Ile Val Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu Gln Lys
 205 210 215

cag atc acc aag atc cag aac ttc cgt gtg tac tac cgg gac tcc cgg 725
 Gln Ile Thr Lys Ile Gln Asn Phe Arg Val Tyr Tyr Arg Asp Ser Arg
 220 225 230

gac cct gtg tgg aag ggc cct gcc aag ctg ctg tgg aag ggc gag ggc 773
 Asp Pro Val Trp Lys Gly Pro Ala Lys Leu Leu Trp Lys Gly Glu Gly
 235 240 245

gcc gtg gtc att cag gac aac tct gac atc aag gtt gtg ccc agg cgc 821
 Ala Val Val Ile Gln Asp Asn Ser Asp Ile Lys Val Val Pro Arg Arg
 250 255 260 265

aag gcc aag att atc cgg gac tac ggc aag cag atg gct ggc gac gac 869
 Lys Ala Lys Ile Ile Arg Asp Tyr Gly Lys Gln Met Ala Gly Asp Asp
 270 275 280

tgt gtg gcc tct cgt caa gat gag gac taa gtccaactac taaactgggg 919
 Cys Val Ala Ser Arg Gln Asp Glu Asp
 285 290

gatattatga t 930

<210> 2

<211> 290

<212> PRT

<213> Artificial Sequence

<223> Description of Artificial Sequence:synthetic gene
 expressing HIV integrase

<400> 2

Met Gly Phe Leu Asp Gly Ile Asp Lys Ala Gln Glu Glu His Glu Lys
 1 5 10 15
 Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro
 20 25 30
 Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys
 35 40 45
 Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln
 50 55 60
 Leu Asp Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala Val His
 65 70 75 80
 Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly
 85 90 95

Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val
 100 105 110
 Lys Thr Val His Thr Asp Asn Gly Ser Asn Phe Thr Ser Thr Thr Val
 115 120 125
 Lys Ala Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu Phe Gly Ile Pro
 130 135 140
 Tyr Asn Pro Gln Ser Gln Gly Val Ile Glu Ser Met Asn Lys Glu Leu
 145 150 155 160
 Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr
 165 170 175
 Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly
 180 185 190
 Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Val Asp Ile Ile Ala Thr
 195 200 205
 Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr Lys Ile Gln Asn
 210 215 220
 Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asp Pro Val Trp Lys Gly Pro
 225 230 235 240
 Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn
 245 250 255
 Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp
 260 265 270
 Tyr Gly Lys Gln Met Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Asp
 275 280 285
 Glu Asp
 290